

Conceptual Index Terms As Used in the Eigen's Political & Historical Quotations Data Base.

By
Lewis D. Eigen
BEA Enterprises Inc.

July 27, 2010

Conceptual Index Terms

This collection has been conceptually indexed. It is the first publicly available data base that does not rely solely on digital, key word indexing.

Difference Between Key Word and Conceptual Indexing

IBM, in the early 1960's, developed the fundamental key word searching mechanisms that are used for searching throughout most of the World Wide Web and most other computer applications. However, it's essential weakness then still remains today. Key words have their limitations, even in context. And the shorter the unit of text that is to be indexed, the more limited the use of key words becomes. Quotations, for the most part, are short. They are particularly susceptible to the "false negative" searches when key word systems are relied upon.

A dramatic example can be seen with the famous quotation of President John F. Kennedy,

***"Say not what your country can do for you.
Say what you can do for your country."***

This quotation clearly has to do with patriotism, duty, and self-interest. But it is extremely difficult for a computer to infer that today and a typical web search for any of those three concepts will not find this quotation. Had a larger work analyzing that quotation been the subject of the word-oriented search, the computer would likely find those three concepts as words in the discussion. But the quotation alone, especially many of the great quotations, will usually not be found if one searches typical systems for the concept terms. Our solution, is a our own unique combination of high and low tech, working in tandem.

In *Eigen's Political & Historical Quotations* you are able to search for quotations involving concepts by searching for the concept terms **without regard for whether or not the actual words of the concepts are utilized in the quotation at all.** It depends not even on a synonym being in the quotation. We have examined each and every quotation and included into the system conceptual analyses - - performed by human beings -- of each and every quotation. In a web analogy, the indexing is more like that of *Yahoo!*® rather than the approach of *Alta Vista*.® This is the main reason that *Eigen's Political & Historical Quotations* has taken so long to develop.

We have examined and tried a wide variety of techniques to avoid the professional, human judgment that goes into this type of conceptual indexing. Thesauruses, artificial intelligence, semantic structures—all have proven very unsatisfactory compared to the old fashioned human indexing. The great computer scientist Alan Turing suggested a test of artificial intelligence whereby the computer succeeds if it can “fool” most people who can’t tell the difference between the interaction with the computer and the interaction with a human (assuming that both are mediated through a display terminal). This collection perhaps provides a special case baseline of Turing’s principle. A true artificial intelligence indexing break-through will be achieved when a computer program can index these quotations so that the user could not tell the difference between our indexing and the computer program’s. This Eigen

Test will doubtless be passed in time. But until that day, probably at least a decade away, we have done the job by human hand and mind.

Subjective Nature of Key Terms

The subjective nature of this form of human indexing is obvious. Reliability is not high among different indexers. However, it is far better than no indexing at all.

One particularly difficult aspect of any Key Term Indexing is the choice of terms. The last decade especially has seen the creation of new terms which are pejorative and emotionally charged. Politicians of various persuasions have tried to get their preferred term used as it carried a political message each time that it is uttered or written. A recent example can be seen with the debate on the “estate tax.” That term has been in legal and political use for decades. In the 1990s, opponents of the estate tax began calling it the “death tax.” The question for us was whether to use the “death tax” term as a conceptual index term or use the more traditional “estate tax.” The rule we have adopted is **NOT** to use contemporary pejorative terms. We have used many pejorative terms if they have stood the test of time and are still in common use decades later. However, unless these meet such a time test we do not use them as conceptual index terms. So we index under “estate tax” and not “death tax.”

Sexist Conceptual Terms

Sexism has to a great extent influenced the development and use of the English language. For each author or editor, a decision must be reached whether to perpetuate the use of sexist language and communicate more facilely or make a serious effort to avoid the sins of the past. We have chosen the former in selecting our indexing terminology. For us, communication trumps the desire to reform the language. So we apologize to those who would have made the alternative decision, and do admit that we are perpetuating a type of wrong, but with noble motives we do use such sexist conceptual indexing terms as:

founding fathers

great men

chairmen

African Americans

Throughout English and American history a variety of different nouns have been used. These have included: "Blacks," "Negroes," and "African Americans." In this collection these terms as conceptual index terms are used interchangeably. Searching for one term should in most cases produce conceptual terms related to the other -- this done with our specially developed Seachonym™ thesaurus software. However, it is suggested that if you do not find one term, try the others as the choice of nouns of the quotees may not correspond to contemporary use. "Blacks" appears to be the most frequent term used. We do not, however include derogatory ethnic terms as seachonyms™. Terms such as "nigger" or "spade" are not index terms or seachonyms™ but the can be searched for and will be found if the author of the quotation actually used the term. We do not edit quotations for issues of taste, sensitivity, or social acceptability.

Black authors and quotees per se however *cannot be determined or searched for per se*. We have not, in most cases, identified the racial identity of the quotees. So, economist Thomas Sowell is simply listed an economist and not a "Black economist." This decision was a difficult one to make. Black and other minority contributions are not appreciated by many who every day benefit from their works, ideas, and quotations. The beneficiary frequently has no idea of the race or ethnicity of the contributor. However, identifying each individual by his or her, racial, religious, and national origins seemed impractical and picking and choosing which ethnicities to identify would end up attempting to resolve an interminable dilemma. Also, the issue of identifying George Washington as "White First President of the United States" seems awkward at best and at worst perhaps will grossly magnify the perception of uniquely white contribution to our culture and ideas.

However where there was an ethnically oriented achievement, an exception was made such as:

Solomon Tshekeisho Plaatje, 1876-1932
South African Journalist & First Black South African Novelist

This principle has been used for other ethnic groups as well.

Hispanic-Americans

As with African Americans, there have been a variety of nouns used to describe the group. "Latinos" and "Hispanics" are the most frequently used and are herein used interchangeably. A parallel set of policies to those for Blacks are followed as has been done for smaller minority groups.

Searching for Nationality

Many quotations refer to nationalities -- Irishmen, Brazilians, Germans, etc. However, under the Conceptual Index Terms, all these nationalities are listed under the country name -- not the nationality. So the search would be for "Ireland," "Brazil," or "Germany." Quotations from non-American authors are usually also indexed under the country of nationality of the quotee except where the quotation is of a thoroughly universal nature. See [Nationality of Quotees](#).

There are several exceptions we make to this general rule. Specifically, when there is a people who either do not have a national homeland or for much of the history of the 20th and 21st centuries did not have a homeland or were scattered among several countries, we use the people appellation rather than the country. The result is that Arabs and Kurds are conceptually indexed as such as are Jews. The Irish became a close call for us. They have so successfully spread throughout the Western world, that the "Irish" are a recognized term and subset of people in virtually all English-speaking countries. And there are many more Irish outside of Ireland than in. Ireland is not even one of the largest population of Irishmen and women any more when compared to Canada and the United States. However, Ireland has

always existed in modern history, so we conceptually index under "Ireland" rather than "Irish."

Turks also had a very different history. Starting as a people with no specific homeland, they populated, conquered and created what is now a homeland, Turkey. Therefore we do not index under the people, "Turks."

England and Great Britain

No single area of conceptual indexing was more difficult than the issue of using the terms "Great Britain," and "England." When the former became the latter is in a little dispute (Scotland and England merged in 1707, and recently Scotland has "devolved" and has a separate Parliament again), but the searcher has little way of determining for which to search. Should one search for "Great Britain AND war" or "England AND war" to find quotations on the English at war? We have made a capricious and arbitrary decision which will make things much easier for the searcher, but doubtless will offend some purists. Nonetheless: "England" is the conceptual term that has been used throughout this collection. If you want to search for British quotations or quotations about Great Britain, use the term "England." While this will also seem offensive to Welsh, Scotch, and Irish, please accept our apologies, but the English have mucked about everyone else's territory for so long, and absorbed so many non-English into their political culture, that no rational classification is possible. When citations are given, and the quotee is known to be Scotch, Irish or Welsh, Indian, etc., the person is usually so identified in the citation.

Even this general rule is not always followed. An example of why can be found with quotations of

Arthur Wellesley, 1769-1852
[1st Duke of Wellington]
British Chief General at Waterloo
Ambassador to France
Prime Minister (Tory)

The famous Duke of Wellington was actually Irish, born in Dublin. He is the only Irish Prime Minister of Great Britain. However, he was never proud of being Irish and never identified with Ireland. Nor do most Irish particularly want to claim him. So we do **not** designate him as specifically Irish.

Also, if the quotation deals with issues of those nations with Great Britain, conceptual terms appear for both the nation and England. So if one is looking for quotations about the relationship between say, Germany and Great Britain, search for "Germany AND England," and not "Germany AND Great Britain." The only use of "Great Britain" as a conceptual term would be a quotation having to do with a significant event of its official creation such as the statement of the Earl of Seafield in the Scottish Parliament as it voted itself out of existence in 1707.

Russia and The Soviet Union

A similar decision was made with respect to the Soviet Union. Conceptual indexing used "Russia" for both the Russian Republic and the Soviet Union. So generally use "Russia" as your search term. We say "generally" because the breakup of the Soviet Union occurred in the midst of the creation of this collection. Just as with England we were making no attempt to make a statement about the conquest of the other Soviet republics and simply used "Russia" for them all. However after the breakup, we were faced with a set of independent nations. This provided a great dilemma. Were we to now describe Joseph Stalin as a Georgian rather than a Russian? We decided in the negative and so Georgian politicians and others are described as "Russian" if they lived and wrote before the breakup; "Georgian," if after. Similarly, for the other republics.

Colonial America

We have created the conceptual indexing term "Colonial America." However, in the indexing process, we had to make some arbitrary decisions based on the notion of at what point does "Colonial America" end. Some argue for the Declaration of Independence being the cut. Others, for the British surrender at Yorktown. An

argument can be made for the ratification of the Constitution, while others would rather the period continue until the first Government sat in New York or when the Bill of Rights was ratified. We found none of these totally satisfactory with different definitions preferable depending on the particular quotation and its context. Consider a quotation made by one of the founding fathers in the 1790s referring to the lack of revolutionary and idealistic spirit that had set in. The comparison persuaded us to designate the quotation as referring to Colonial America even though the date was later than reasonable cut-offs that would be implied by different definitions.

So we chose not to be consistent in the use of the index term, "Colonial America," and used our editorial judgment, but with few Colonial America indexed which were uttered after 1790.

END